TECHNICAL REPORT
R-411
March 2013

Comments on "Surrogates measures and consistent surrogates" (by Tyler VanderWeele)

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
(310) 825-3243
judea@cs.ucla.edu

I commend Professor VanderWeele for providing a lucid description of the "surrogate paradox" and, through it, a comprehensive discussion of the current state of thinking about surrogate endpoints, their function in experimental studies, and the various approaches devised to give them formal underpinnings.

The first question that came to mind in reading VanderWeele's paper was: can we explain the phenomenon in simple terms, divorced from the technical vocabulary that was devised to formulate notions such as "indirect effect," "principled strata," "proportion-mediated," and perhaps others? My second question was: If we take the negation of the "surrogate paradox" as a criterion for "good" surrogate, why can't we create a new, formal definition of "surrogacy" that (1) will automatically avoid the paradox and (2) will settle, once for all, the disputes (among theoreticians) as to what "approach" is best for defining surrogates (Joffe and Green, 2009, pp. 530–538; Pearl, 2011).

In thinking about these two questions, I came across a simple way of explaining how the paradox comes about and, indirectly, why the requirement of avoiding the paradox could not, in itself, constitute a satisfactory definition of surrogacy.

As with other paradoxes of causal inference (e.g., Simpson's paradox, Berkson's paradox, suppression effect, reverse regression) a good starting point is linear models, where the emergence of "paradoxical" phenomena can be examined under the powerful "microscope" of path analysis and elementary linear regression (Pearl, 2013a). If a paradox emerges in linear models, we can be sure that its origin does not rest with effect heterogeneity or idiosyncratic non-linearities, but with the age-old confusion between regression and causation (Pearl, 2013b).

Indeed, starting with the simple linear model of Fig. 1(a), we can write the effects of A on S and on Y, as well as the correlation between the S and Y in terms of the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Comments on 'Surrogates measures and consistent surrogates' (by Tyler VanderWeele)	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Los Angeles,Computer Science Department,Los Angeles,CA,90095-1596**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**4**	

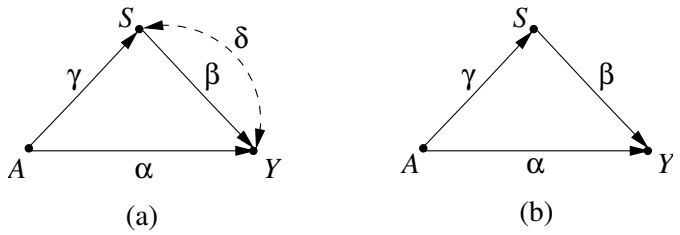

Figure 1: Path diagram in which S acts as a surrogate for the effect of A on Y, demonstrating the "surrogate paradox" under both confounded (a) and unconfounded (b) models.

structural parameters α, β, γ and δ.[1]

$$E(Y_1 - Y_0) = \alpha + \beta\gamma \tag{1}$$

$$E(Y|S = 1, A = a) - E(Y|S = 0, A = a) = \beta + \delta \tag{2}$$

$$E(S_1 - S_0) = \gamma \tag{3}$$

The surrogate paradox will be exhibited when the effect of treatment A on the surrogate S (3) is positive, S and Y are positively correlated (2), but the effect of A on Y is negative (1), that is, when the structural parameters satisfy:

$$\alpha + \beta\gamma < 0 \tag{4}$$

$$\beta + \delta > 0 \tag{5}$$

$$\gamma > 0 \tag{6}$$

Clearly, for any $\gamma > 0$ and any β, one can find α sufficiently negative and δ sufficiently positive so as to satisfy (4) and (5). Moreover, even for the unconfounded case, $\delta = 0$, shown in Fig. 1(b), the three inequality can be satisfied with $\beta > 0$ and α sufficiently negative, namely, $\alpha < -\beta\gamma$.

We conclude that the surrogate paradox may occur in very common models; it does not require confounding, nor interaction or heterogeneity. It requires only that the direct effect of A on Y be sufficiently negative for the paradox to surface. This of course is an unlikely situation in practice. A treatment that has such a negative direct effect on outcome would rarely be a candidate for surrogacy analysis. In practice, the paradox is more likely to take place under confounding conditions ($\delta > 0$) where even a positive α and a negative β will permit it to surface.

We now address the question of why we cannot pose the avoidance of the surrogate paradox, namely, the positivity of all quantities on the left hand side of Eqs. (1)–(3) as a formal definition of a "good" surrogate. Indeed, unlike Simpson's paradox, which stems from a misinterpretation of statistical data (Pearl, 2009, Ch. 6), negating the surrogate paradox expresses precisely what we expect a "good" surrogate to do. It is expected to provide a good prediction of outcome, once it is found to be positively

[1]We assume a randomized trial, hence, A and S are not confounded nor are A and Y. δ stands for the covariance of the "disturbances" affecting S and Y.

affected by the treatment. Why, then, have researchers labored to define "good" surrogates using fancy formalisms such as "indirect effect," "principal strata,[2]" or "proportion-mediated" (Joffe and Green, 2009) instead of constraining Eqs. (1)–(3) with the proper inequalities?

The reason, I believe, is that definitions are expected to be formulated in terms of the knowledge available to the investigator at the time of the study, and this knowledge consists of qualitative understanding of the model's structure prior to seeing the data, or quantitative assessments of the parameters after examining the data. Eqs. (4)–(6) show that structural knowledge is not sufficient to protect us from the paradox. The paradox may surface even when $\alpha = 0$ (strong surrogacy) or $\beta = 0$. About the only structural condition to prevent the paradox is $\alpha = \delta = 0$, which amounts to perfect mediation (Prentice, 1989). As to quantitative protection from the paradox, the confounding model of Fig. 1 does not permit the identification of α, β, and δ, or, in the nonparametric case, of direct and indirect effects.

Another important consideration is robustness. Pearl and Bareinboim (2011) argued that good prediction of the effect of A on Y should not be the sole criterion for judging surrogacy, but must be accompanied with a requirement of robustness. Let us imagine two studies. In the first, we measure the effects of A on both S and Y and confirm that S is a good surrogate, that is, knowing the effect of treatment on S allows prediction of the effect of treatment on the outcome. Once S is proclaimed a "surrogate," it invites efforts to find direct means of controlling S. For example, if cholesterol level (S) is found to be a predictor of heart disease in a long run (Y), drug manufacturers would rush to offer cholesterol-reducing substances for public consumption. As a result, both the prior $P(S = s)$ and the treatment-dependent probability $P(S = s|A = a)$ would undergo a change. For S to be a good surrogate, we should be able to re-assess the effect of the treatment $E(Y_1 - Y_0)$ in a new population in which the effect of treatment on S has changed, and in which access to Y is no longer available. Instead, we have an experiment to assess the new value of $E(S_1 - S_0)$. Pearl and Bareinboim (2011) have shown that, if we assume that the disparity between the two populations lies only in the difference in $E(S_1 - S_0)$ (the surrogate's susceptibility to treatment) the effect of treatment on the outcome under the new conditions can still be estimated from the two studies, provided S and Y are not confounded.

Acknowledgment

I thank the editor for the opportunity to comment on this important paper.

This research was supported in parts by grants from NSF #IIS-1249822 and ONR #N00014-13-1-0153 and #N00014-10-1-0933.

[2]The choice of "principal strata" to define surrogacy is particularly inadequate, for these strata are empty in the case of continuous S (Pearl, 2011).

References

Joffe, M. and Green, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65** 530–538.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2011). Principal stratification a goal or a tool? *The International Journal of Biostatistics* **7**. Article 20, DOI: 10.2202/1557-4679.1322. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r382.pdf>.

Pearl, J. (2013a). Linear models: A useful "microscope" for causal analysis. Tech. Rep. R-409, <http://ftp.cs.ucla.edu/pub/stat_ser/r409.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Forthcoming, *Journal of Causal Inference*.

Pearl, J. (2013b). Trygve Haavelmo and the emergence of causal calculus. Tech. Rep. R-391, <http://ftp.cs.ucla.edu/pub/stat_ser/r391.pdf>, University of California Los Angeles, Computer Science Department, CA. Forthcoming, *Econometric Theory*, special issue on Haavelmo Centennial.

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-11)*. Menlo Park, CA. Available at: <http://ftp.cs.ucla.edu/pub/stat_ser/r372a.pdf>.

Prentice, R. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* **8** 431–440.